Investigation of the Relationship between Diesel Fuel Properties and Emissions from Engines with Fuzzy Linear Regression

Yuanwang Deng

College of mechanical and vehicle engineering, Hunan University, Changsha, 410082, People's Republic of China

dengyuanwang610@126.com

Abstract

Selecting the fuel properties as the independent variables, and the emissions from engines as the dependent variables, the fuzzy linear regression equations are constructed in this paper with the fuzzy linear regression; as well as the simulation and prediction results calculated with the equations are analyzed. Satisfactory results are achieved by selecting an appropriate degree of membership. The effect of the degree of the membership on the fuzzy regression coefficients is also analyzed. There was no effect on the fuzzy centers of the fuzzy regression coefficients of the equations for total hydrocarbons (HC), carbon monoxide (CO), nitrogen oxides (NO_x). There was a little effect on those of the equation for particulates (PM). The fuzzy widths of the fuzzy regression coefficients increase when the degree of the membership increases. When the biases of the center values of the simulation results to the actual values are big or the intervals of the simulation results are too big to have meaning, additional restrictions should be used to improve the simulation accuracy of the equations.

Keywords

Emissions from Engines; Fuel Properties; Fuzzy Linear Regression; Particulates

Introduction

The air pollution of the cities in China has become more and more serious due to automotive emissions in the last few years, so new national standards for automotive emissions were enacted by the national environmental protective agency on March 10th, 1999 and put into effect enacted on January 1st, 2000. The standards require that the levels of the automotive emissions must reach European EURO-Ⅰemissions standards, thus the manufacturers must introduce new techniques to control the emissions. Since there are no special vehicle diesel fuels in China at present, their standards mechanically apply the standards of the light diesel fuels, therefore their quality cannot reach the demand of the legislation. Investigation of the effect of the fuel properties on the emissions from engines is important for controlling the emissions from engines and thus decreasing the air pollution in our country.

Based on the test, selecting the fuel properties as the independent variables and the emissions from engines as the dependent variables, linear regression equations can be constructed with statistical method. The relationship between the fuel properties and the emissions can be analyzed and the emissions can be predicted with the equations. Since the fuel properties are correlative with each other, when studying the effect of one of the properties on the emissions, the property must be separated from the other properties. The relationship between the fuel property and the emissions is fuzzy. This paper constructs the fuzzy linear regression equations on the relationship between the fuel properties and the emissions with fuzzy mathematics theory. The results obtained with the equations are analyzed and discussed.

Fuzzy Linear Regression

Assuming that $x_1, x_2, \cdots, x_n$ are the independent variables and y is the dependent variable, the samples $x_{i1}, x_{i2}, \cdots x_{in}$ — $y_i, (i = 1,2,\cdots,m)$ are known. According to the classical linear regression, the multivariate linear regression model is denoted (Hu G et al., 1990):

$$y = c_0 + c_1 x_1 + c_2 x_2 + \cdots + c_n x_n + \varepsilon , \tag{1}$$

Where, $c_0, c_1, c_2, \cdots, c_n$ are parameters to be estimated, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m)$ is an unobserved random error, satisfying $\varepsilon \sim N(0,\sigma^2)$.

The classical linear regression considers that the error ε between the actual y and the calculated y^* of the

dependent variable is caused by the error in the test. The sum of squares of the errors should be minimized when constructing the linear regression equation with the known samples. Thus the multivariate linear regression equation is obtained:

$$y^* = \hat{c}_0 + \hat{c}_1 x_1 + \hat{c}_2 + \cdots + \hat{c}_n x_n, \quad (2)$$

Here, $\hat{c}_0, \hat{c}_1, \hat{c}_2, \cdots, \hat{c}_n$ are the estimated regression coefficients.

The fuzzy linear regression was founded in 1980s (Hideo Tanaka et al., 1982) and has been applied in engine engineering (Li B et al., 2000; Wu J et al., 1999). It considers that the error ε between the actual y and the calculated y^* of the dependent variable is caused by the fuzziness of the system. The following fuzzy linear regression equation is obtained with the fuzzy numbers instead of the coefficients in the classical linear regression equation:

$$\underset{\sim}{y}^* = \underset{\sim}{A_0}^* + \underset{\sim}{A_1}^* x_1 + \underset{\sim}{A_2}^* x_2 + \cdots + \underset{\sim}{A_n}^* x_n, \quad (3)$$

Where, $\underset{\sim}{A_i}^*, i = 0,1,2,\cdots,n$ are the estimated fuzzy regression coefficients, usually adopting the symmetrical triangular fuzzy number, whose general form is: $\underset{\sim}{A_i}^* = (\xi_i, \eta_i), i = 0,1,2,\cdots,n$. Where, ξ_i is the fuzzy center, η_i （$\eta_i \ge 0$）is the fuzzy width, and its membership function is defined:

$$\mu_{\underset{\sim}{A_i}^*}(a_i) = \begin{cases} 1 - \dfrac{|a_i - \xi_i|}{\eta_i} & \xi_i - \eta_i \le a_i \le \xi_i + \eta_i \\ 0 & otherwise \end{cases}. \quad (4)$$

Assuming $x^T = (1, x_1, x_2, \cdots, x_n)$, $|x|^T = (1, |x_1|, |x_2|, \cdots, |x_n|)$, $\xi^T = (\xi_0, \xi_1, \cdots, \xi_n)$, $\eta^T = (\eta_0, \eta_1, \cdots, \eta_n)$, according to the arithmetic operations of the fuzzy number (Li H et al., 1994), $\underset{\sim}{y}^* = (x^T \xi, \eta^T |x|)$ is obtained. Its membership function is:

$$\mu_{\underset{\sim}{y}^*}(y) = \begin{cases} 1 - \dfrac{|y - x^T \xi|}{\eta^T |x|}, & x \ne 0 \\ 1, & x = 0, y = 0 \\ 0, & x = 0, y \ne 0 \end{cases}. \quad (5)$$

We aim to obtain the fuzzy numbers: $\underset{\sim}{A_i}^*, i = 0,1,2,\cdots,n$, now let's see the calculation method.

According to the above-mentioned, $\mu_{\underset{\sim}{y}^*}(y)$ must be decided by the decision-maker, subjecting to $\mu_{\underset{\sim}{y}^*}(y) \ge H$, where H ($H \in [0,1]$) is the degree of the membership of the fuzzy linear regression model selected by the decision-maker. The fuzzy width of $\underset{\sim}{y}^*$ must be minimized to reduce the vagueness of the result. So the above fuzzy linear regression problem is transformed into the solution of the following linear programming problem:

$$J = \min(\eta_0 + \eta_1 + \cdots + \eta_n), \eta_j \ge 0, j = 0,1,\cdots,n \quad (6)$$

Subjecting to:

$$1 - \frac{|y_i - x_i^T \xi|}{\sum_j \eta_j x_{ij}} \ge H, i = 1,2,\cdots,m; j = 0,1,\cdots,n \cdot \quad (7)$$

That is to say:

$$\begin{cases} J = \min(\eta_0 + \eta_1 + \cdots + \eta_n), \eta_j \ge 0, j = 0,1,\cdots,n \\ s.t. \\ (1-H)\sum_j \eta_j x_{ij} + x_i^T \xi \ge y_i \\ (1-H)\sum_j \eta_j x_{ij} - x_i^T \xi \ge -y_i \\ i = 1,2,\cdots,m; j = 0,1,\cdots,n \end{cases}. \quad (8)$$

The fuzzy regression coefficients $\underset{\sim}{A_i}^* = (\xi_i, \eta_i), i = 0,1,2,\cdots,n$ are obtained by solving the above linear programming problem.

When using the fuzzy linear regression to solve the actual problem, we may also meet another type of the samples: $x_{i1}, x_{i2}, \cdots x_{in}$ — $\underset{\sim}{y_i} = (y_i, e_i)$, $i = 1,2,\cdots,m$, that is to say, the outputs of the samples are fuzzy numbers. Here, we also consider them as the symmetrical triangular fuzzy numbers. The degree of the membership of the calculated $\underset{\sim}{y_i}^* (i = 1,2,\cdots,m)$ to the actual $\underset{\sim}{y_i}$ is denoted by $\bar{h}_i$, $\bar{h}_i = max(h)$, h satisfies $\underset{\sim}{y_i}^h \subset \underset{\sim}{y_i}^{*h}$. Here:

$$\underset{\sim}{y}_i^{*h} = \left\{ y \middle| \mu_{\underset{\sim}{y}_i^*}(y) \ge h \right\} \text{ and } \underset{\sim}{y}_i^{h} = \left\{ y \middle| \mu_{\underset{\sim}{y}_i}(y) \ge h \right\} \quad (9)$$

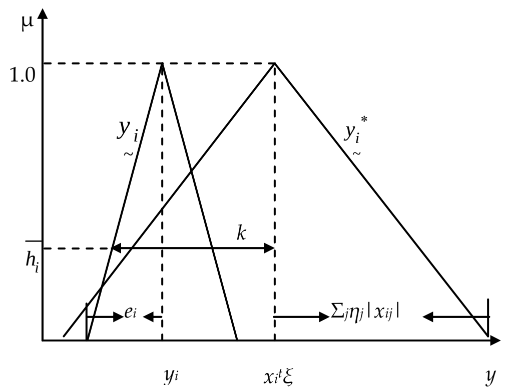

FIG.1 THE DEGREE OF THE MEMBERSHIP OF $\underset{\sim}{y}_i^*$ TO $\underset{\sim}{y}_i$

Figure 1 gives the degree of the membership of $\underset{\sim}{y}_i^*$ to the actual $\underset{\sim}{y}_i$. The following formula is obtained with the similarity of the right triangles:

$$1:(1-\overline{h}_i) = (\sum_j \eta_j |x_{ij}|):k,$$

Where, $k = |y_i - x_i^t\xi| + e_i(1-\overline{h}_i)$.

Hence

$$\overline{h}_i = 1 - \frac{|y_i - x_i^t\xi|}{\sum_j \eta_j |x_{ij}| - e_i}. \quad (10)$$

As mentioned above, the fuzzy linear regression problem in this condition can also be transformed into the following linear programming problem:

$$J = \min(\eta_0 + \eta_1 + \cdots + \eta_n), \eta_j \ge 0, j = 0,1,\cdots,n \quad (11)$$

Subjecting to:

$$1 - \frac{|y_i - x_i^t\xi|}{\sum_j \eta_j |x_{ij}| - e_i} \ge H, i = 1,2,\cdots,m; j = 0,1,\cdots,n \quad (12)$$

That is to say:

$$\begin{cases} J = \min(\eta_0 + \eta_1 + \cdots + \eta_n), \eta_j \ge 0, j = 0,1,\cdots,n \\ s.t. \\ (1-H)\sum_j \eta_j x_{ij} + x_i^T\xi \ge y_i + (1-H)e_i \\ (1-H)\sum_j \eta_j x_{ij} - x_i^T\xi \ge -y_i + (1-H)e_i \\ i = 1,2,\cdots,m; j = 0,1,\cdots,n \end{cases} \quad (13)$$

The fuzzy regression coefficients are also obtained by the solution of the above linear programming problem. Comparing the membership functions $\mu_{\underset{\sim}{y}^*}(y)$ of the two types of the samples, it is found that, when $e_i = 0$, the membership function $\mu_{\underset{\sim}{y}^*}(y)$ of the latter is the same as that of the former, so the former is the special example of the latter.

Application and Analysis

The samples data used in this paper come from the paper (M.Hublin et al., 1996). Eleven fuels with different properties were used in that paper to do the tests with several light-duty diesel engines. Table 1 gives the properties of the fuels. Table 2 gives the emissions from engines HC, CO, NO_x, PM. By analyzing the properties of the fuels, fuels 1, 3, 4, 5, 7, 8, 9, 11, along with the corresponding results of the emissions, are selected as the simulation samples 1-8. They are used to construct the fuzzy linear regression equations. Fuels 2, 6, 10, along with the corresponding results of the emissions, are selected as the prediction samples 1-3. They are used to analyze the prediction ability of the fuzzy linear regression equations.

TABLE 1 THE PROPERTIES OF THE FUELS

No.	1	2	3	4	5	6	7	8	9	10	11
ρ/ kg/m³	829.2	828.8	857.0	855.1	828.8	855.5	826.9	855.1	855.4	826.6	827
p /%	1.0	7.7	1.1	7.4	7.1	7.6	1	7.3	8	1.1	0.9
CN /r	51.0	50.2	50.0	50.3	50.6	50.2	49.5	54.8	59.1	58	57.1
T_{95}/℃	344	349	348	344	346	371	326	345	344	347	329

No.—fuel code, ρ— density, p—polyaromatics, CN—cetane number, T_{95}—back end distillation

TABLE 2 THE EMISSIONS FROM ENGINES /g/km

No.	1	2	3	4	5	6	7	8	9	10	11
HC	0.085	0.085	0.111	0.103	0.085	0.099	0.089	0.083	0.073	0.063	0.066
CO	0.432	0.431	0.551	0.517	0.446	0.513	0.454	0.434	0.378	0.331	0.327
NO_x	0.556	0.564	0.532	0.546	0.556	0.539	0.554	0.552	0.559	0.534	0.551
PM	0.048	0.052	0.061	0.063	0.051	0.066	0.045	0.065	0.066	0.051	0.049

Selecting density, polyaromatics, cetane number, and back end distillation as the independent variables, and one of the emissions HC, CO, NO_x, or PM as the dependent variable, according to (3), the following fuzzy linear regression equation is obtained:

$$\underset{\sim}{y}^* = \underset{\sim}{A_0}^* + \underset{\sim}{A_1}^* x_1 + \underset{\sim}{A_2}^* x_2 + \underset{\sim}{A_3}^* + \underset{\sim}{A_4}^* x_4 \,. \tag{14}$$

According to (8), the solution of (14) is transformed into the solution of the following linear programming problem with the simulation samples1-8:

$$\begin{cases} J = min(\eta_0 + \eta_1 + \cdots + \eta_4), \eta_j \geq 0, j = 0,1,\cdots,4 \\ s.t. \\ (1-H)\sum_j \eta_j x_{ij} + x_i^T \xi \geq y_i \\ (1-H)\sum_j \eta_j x_{ij} - x_i^T \xi \geq -y_i \\ i = 1,2,\cdots,8; j = 0,1,\cdots,4 \end{cases} . \tag{15}$$

Because the sizes of different independent variables vary widely, in order to remove the effect of the sizes of them on the regression coefficients, the data is transformed into deviations from the average. Selecting H=0.8, the coefficients are calculated with the simplex method (He J et al., 1985), and shown in table 3. For example, for HC, using the coefficients in table 3, the fuzzy linear regression equation is denoted with the following formula:

$$\begin{aligned} y = & (0.0862129618, 0.0103603254) + (6.44910839659 \times 10^{-4}, 8.21203534238 \times 10^{-5}) x_1 \\ & + (-8.63466678222 \times 10^{-4}, 0) x_2 + (-0.0033727637, 0) x_3 \\ & + (2.57416992700 \times 10^{-5}, 2.36034617082 \times 10^{-4}) x_4 \end{aligned}$$

The equations of CO, NO_x, and PM can also be denoted with a similar formula using the coefficients in table 3. Comparing with table 2, because the data of each independent variable has been transformed into deviations from the average, the average of the dependent variable can be explained by the size of the fuzzy center of the coefficient for the constant x_0. The size of the fuzzy center of each coefficient is greatly different from its fuzzy width, except for the size of the fuzzy centers of the coefficients of the independent variable x_4 of the equations of HC and CO which are an order of magnitude smaller than their fuzzy widths. The size of the fuzzy center of the coefficient of the independent variable x_3 of the equation of PM is the same order of magnitude as its fuzzy width. The sizes of the fuzzy centers of other coefficients of the equations are an order of magnitude bigger than their fuzzy widths. This shows that the calculated coefficients are rational. It is also found from table 3 that the fuzzy widths of some coefficients are equal to zero, in this case the coefficients essentially become accurate numbers. There are some coefficients whose fuzzy widths are not equal to zero, so the results obtained with these equations are still fuzzy numbers. When all the coefficients become accurate numbers, the equations will become the classical linear regression equations, and the results will also become accurate numbers.

TABLE 3 THE FUZZY REGRESSION COEFFICIENTS OF H=0.8

		x_0	x_1	x_2	x_3	x_4
HC	ξ	0.0862129618	$6.44910839659 \times 10^{-4}$	$-8.63466678222 \times 10^{-4}$	-0.0033727637	$2.57416992700 \times 10^{-5}$
	η	0.0103603254	$8.21203534238 \times 10^{-5}$	0	0	$2.36034617082 \times 10^{-4}$
CO	ξ	0.4411455297	0.0029706148	-0.0014849168	-0.0170481746	$2.64004752996 \times 10^{-4}$
	η	0.0343943595	$1.20238584005 \times 10^{-4}$	0	0	0.0022102577
NO_x	ξ	0.5506190827	$-5.26722941176 \times 10^{-4}$	0.0014825633	$7.92863929877 \times 10^{-4}$	$1.12342603883 \times 10^{-4}$
	η	0.0229046626	0	0	0	0
PM	ξ	0.0559301482	$4.64879738324 \times 10^{-4}$	$4.25842266960 \times 10^{-4}$	$4.25468163699 \times 10^{-4}$	$7.76985395646 \times 10^{-5}$
	η	0.0014994813	0	0	$3.53882164757 \times 10^{-4}$	0

TABLE 4 THE SIMULATION RESULTS OF H=0.8 CALCULATED WITH THE FUZZY LINEAR REGRESSION EQUATIONS

The simulation samples		1	2	3	4	5	6	7	8
HC	Upper limits	0.0921	0.1150	0.1067	0.0882	0.0930	0.0917	0.0767	0.0679
	Centers	0.0870	0.1083	0.1006	0.0829	0.0901	0.0855	0.0706	0.0647
	Lower limits	0.0820	0.1017	0.0944	0.0776	0.0873	0.0793	0.0644	0.0615
CO	Upper limits	0.4600	0.5667	0.5410	0.4593	0.4540	0.4658	0.3910	0.3290
	Centers	0.4400	0.5406	0.5194	0.4371	0.4540	0.4431	0.3694	0.3257
	Lower limits	0.4200	0.5144	0.4978	0.4149	0.4540	0.4204	0.3478	0.3224
NO_x	Upper limits	0.5629	0.5480	0.5582	0.5720	0.5609	0.5617	0.5659	0.5670
	Centers	0.5514	0.5366	0.5467	0.5606	0.5494	0.5502	0.5544	0.5556
	Lower limits	0.5400	0.5251	0.5353	0.5491	0.5380	0.5388	0.5430	0.5441
PM	Upper limits	0.0486	0.0613	0.0630	0.0509	0.0452	0.0657	0.0687	0.0501
	Centers	0.0482	0.0610	0.0626	0.0506	0.0451	0.0646	0.0668	0.0485
	Lower limits	0.0477	0.0477	0.0623	0.0502	0.0449	0.0635	0.0649	0.0470

Table 4 gives the center values, upper limits and lower limits of the simulation results. It is found from the table that, except for the actual value of PM of sample 4 which is out of the interval of the simulation results, the other actual values lie in the intervals of the simulation results, because the intervals of the simulation results are much smaller than the center values of the simulation results which are rational. This shows that all of the fuzzy linear regression equations can present correctly the functional relationship between the independent variables and the dependent variables. We can use the equations to predict the emissions from engines.

Table 5 gives the prediction results. Comparing with the actual values in table 2, it is found that, except for the actual values of NO_x of sample 3 (fuel 10) and PM of samples 1,2 (fuel 2,6) which are out of the intervals of the prediction results, the other actual values lie in the intervals of the prediction results. When the actual values do not lie in the intervals of the prediction results, they approach the upper limits or the lower limits of the prediction results. Comparing the upper limits with the lower limits of the prediction results, it is also found that, when the actual values are not in the intervals of the prediction results, the intervals of the prediction results are small. This is caused by the small fuzzy width of the system. Hence, in order to improve the prediction ability of the equations, the fuzzy width of the coefficients must be increased.

TABLE 5 THE PREDICTION RESULTS OF H=0.8 ALCULATED WITH THE FUZZY LINEAR REGRESSION EQUATIONS

The prediction samples		1	2	3
HC	Upper limits	0.0894	0.1110	0.0670
	Centers	0.0838	0.1017	0.0617
	Lower limits	0.0782	0.0924	0.0564
CO	Upper limits	0.4694	0.5806	0.3368
	Centers	0.4438	0.5291	0.3136
	Lower limits	0.4183	0.4777	0.2904
NO_x	Upper limits	0.5729	0.5612	0.5703
	Centers	0.5615	0.5498	0.5588
	Lower limits	0.5500	0.5383	0.5474
PM	Upper limits	0.0512	0.0653	0.0519
	Centers	0.0509	0.0650	0.0502
	Lower limits	0.0506	0.0647	0.0486

Further Analysis and Discussion of the Degree of the Membership

The degree of the membership is selected by the decide-maker in terms of the demand of the actual problem. According to (15), H influences the solution of the coefficients of the equations, so it is necessary to further analyze H. Here, H=0.5 and H=0.9 are selected to analyze the effect of them on the coefficients of the equations.

TABLE 6 THE FUZZY REGRESSION COEFFICIENTS OF H=0.5

		x_0	x_1	x_2	x_3	x_4
HC	ξ	0.0862129618	6.44910839659×10^{-4}	-8.63466678222×10^{-4}	-0.0033727637	2.57416992700×10^{-5}
	η	0.0041441302	3.28481413695×10^{-5}	0	0	9.44138468326×10^{-5}
CO	ξ	0.4411455297	0.0029706148	-0.0014849168	-0.0170481746	2.64004752995×10^{-4}
	η	0.0137577438	4.80954336022×10^{-5}	0	0	8.84103095913×10^{-4}
NO_x	ξ	0.5506190827	-5.26722941176×10^{-4}	0.0014825633	7.92863929877×10^{-4}	1.12342603883×10^{-4}
	η	0.0091618651	0	0	0	0
PM	ξ	0.0561270521	4.61977224464×10^{-4}	3.86165856436×10^{-4}	3.95489713119×10^{-4}	7.46933031100×10^{-5}
	η	7.0782075363×10^{-4}	0	0	0	0

Table 6 gives the coefficients of the equations for H=0.5. Comparing with table 3, it is found that the fuzzy centers of the coefficients of the equations of HC, CO and NO_x are the same as those of H=0.8. However, the fuzzy centers of the coefficients of the equations of PM are different from those of H=0.8, but both fuzzy centers are close. Except for the fuzzy widths which are equal to zero, the other fuzzy widths of the coefficients of the equations are smaller than those of H=0.8. So when H=0.5, simulation HC, CO and NO_x with the equations, the center values of the simulation results are not changed. However, simulating PM with the equation, the center values of the simulation results are changed. The intervals of the simulation results are decreased.

Table 7 gives the center values, upper limits and lower limits of the simulation results. It is found from the table that the actual values of HC depart from the center values and lie in the upper or lower limits of the simulation results. For CO, except for sample 3 (fuel 4) where the actual values approach the center value of the simulation result, the other actual values present the same law as those of HC. For NO_x, except for 3, 6 (fuels 4, 8) where the actual values approach the center values of the simulation results, the other actual values also present the same law as those of HC. Comparing the actual values with the simulation results of PM, it is found that the actual values of samples 2, 3, 4, 8（fuels 3, 4, 5, 11）lie in the interval of the simulation results. The other actual values do not lie in the intervals of the simulation results. Hence, when H=0.5, the simulation accuracy of the equations is worse than that of H=0.8.

TABLE 7 THE SIMULATION RESULTS OF H=0.5 CALCULATED WITH THE FUZZY LINEAR REGRESSION EQUATIONS

The simulation samples		1	2	3	4	5	6	7	8
HC	Upper limits	0.0890	0.1110	0.1030	0.0850	0.0913	0.0880	0.0730	0.0660
	Centers	0.0870	0.1083	0.1006	0.0829	0.0901	0.0855	0.0706	0.0647
	Lower limits	0.0850	0.1057	0.0981	0.0808	0.0890	0.0830	0.0681	0.0635
CO	Upper limits	0.4480	0.5510	0.5280	0.4460	0.4540	0.4522	0.3780	0.3270
	Centers	0.4400	0.5406	0.5194	0.4371	0.4540	0.4431	0.3694	0.3257
	Lower limits	0.4320	0.5301	0.5107	0.4282	0.4540	0.4340	0.3607	0.3243
NO_x	Upper limits	0.5560	0.5412	0.5513	0.5652	0.5540	0.5548	0.5590	0.5602
	Centers	0.5514	0.5366	0.5467	0.5606	0.5494	0.5502	0.5544	0.5556
	Lower limits	0.5468	0.5320	0.5421	0.5560	0.5448	0.5457	0.5498	0.5510
PM	Upper limits	0.0489	0.0617	0.0631	0.0511	0.0459	0.0649	0.0670	0.0492
	Centers	0.0486	0.0614	0.0627	0.0507	0.0456	0.0646	0.0666	0.0488
	Lower limits	0.0482	0.0610	0.0624	0.0504	0.0452	0.0642	0.0662	0.0485

TABLE 8 THE PREDICTION RESULTS OF H=0.5 CALCULATED WITH THE FUZZY LINEAR REGRESSION EQUATIONS

The prediction samples		1	2	3
HC	Upper limits	0.0860	0.1054	0.0638
	Centers	0.0838	0.1017	0.0617
	Lower limits	0.0816	0.0979	0.0596
CO	Upper limits	0.4540	0.5497	0.3229
	Centers	0.4438	0.5291	0.3136
	Lower limits	0.4336	0.5085	0.3043
NO_x	Upper limits	0.5661	0.5543	0.5634
	Centers	0.5615	0.5498	0.5588
	Lower limits	0.5569	0.5452	0.5542
PM	Upper limits	0.0514	0.0653	0.0508
	Centers	0.0510	0.0650	0.0504
	Lower limits	0.0507	0.0646	0.0501

Table 8 gives the prediction results. Comparing with the actual values of table 2, it is found that the actual values of HC lie in the intervals of the prediction results. For CO, the actual value of sample 2 (fuel 6) lies in the interval of the prediction result, otherwise the other two actual values do not lie in the intervals of the prediction results. For NO_x, the actual value of sample 1 (fuel 2) lies in the interval of the prediction result, yet the other two actual values exclude to do so. However, for PM, none of the actual values lie in the intervals of the prediction results. Hence, the prediction results are not good.

Table 9 gives the coefficients of the equations for H=0.9. Comparing with table 3, it is found that the fuzzy centers of the coefficients of the equations for HC, CO and NO_x are the same as those of H=0.8. The fuzzy centers of the coefficients of the equation for PM are different from those of H=0.8, except that when the fuzzy widths are equal to zero, the other fuzzy widths are bigger than those of H=0.8. Similarly comparing with table 6, the same conclusion is also reached. Hence, the intervals of the simulation and prediction results of the equations are bigger than those of H=0.8 and H=0.5.

Table 10 gives the center values, upper limits and lower limits of the simulation results. Comparing with the actual values of table 2, it is found that all of the actual values lie in the intervals of the simulation results. Comparing with the simulation results of table 4, the center values of the simulation results of HC, CO and NO_x are same, and the center values of the simulation results of PM are different, but their biases to the actual values are all small. However, the intervals of the simulation results increase. Similarly, comparing with table 7, the same conclusion is also found.

TABLE 9 THE FUZZY REGRESSION COEFFICIENTS OF H=0.9

		x_0	x_1	x_2	x_3	x_4
HC	ξ	0.0862129618	$6.44910839659\times10^{-4}$	$-8.63466678222\times10^{-4}$	-0.0033727637	$2.57416992700\times10^{-5}$
	η	0.0207206508	$1.64240706848\times10^{-4}$	0	0	$4.72069234163\times10^{-4}$
CO	ξ	0.4411455297	0.0029706148	-0.0014849168	-0.0170481746	$2.64004752996\times10^{-4}$
	η	0.0687887191	$2.40477168011\times10^{-4}$	0	0	0.0044205155
NO_x	ξ	0.5506190827	$-5.26722941176\times10^{-4}$	0.0014825633	$7.92863929877\times10^{-4}$	$1.12342603883\times10^{-4}$
	η	0.0458093253	0	0	0	0
PM	ξ	0.0559861280	$4.62631797962\times10^{-4}$	$4.82195589361\times10^{-4}$	$4.01690942831\times10^{-4}$	$8.17946018667\times10^{-5}$
	η	0.0030556324	0	0	$9.25949221079\times10^{-4}$	0

TABLE 10 THE SIMULATION RESULTS OF H=0.9 CALCULATED WITH THE FUZZY LINEAR REGRESSION EQUATIONS

The simulation samples		1	2	3	4	5	6	7	8
HC	Upper limits	0.0971	0.1217	0.1128	0.0934	0.0958	0.0979	0.0828	0.0711
	Centers	0.0870	0.1083	0.1006	0.0829	0.0901	0.0855	0.0706	0.0647
	Lower limits	0.0769	0.0950	0.0883	0.0724	0.0845	0.0730	0.0583	0.0584
CO	Upper limits	0.4801	0.5928	0.5626	0.4815	0.4540	0.4885	0.4126	0.3323
	Centers	0.4400	0.5406	0.5194	0.4371	0.4540	0.4431	0.3694	0.3257
	Lower limits	0.4000	0.4883	0.4762	0.3927	0.4540	0.3977	0.3261	0.3190
NO_x	Upper limits	0.5743	0.5595	0.5696	0.5835	0.5723	0.5731	0.5773	0.5785
	Centers	0.5514	0.5366	0.5467	0.5606	0.5494	0.5502	0.5544	0.5556
	Lower limits	0.5285	0.5137	0.5238	0.5377	0.5265	0.5273	0.5315	0.5327
PM	Upper limits	0.0488	0.0612	0.0633	0.0514	0.0450	0.0672	0.0713	0.0518
	Centers	0.0481	0.0610	0.0629	0.0509	0.0450	0.0648	0.0669	0.0483
	Lower limits	0.0474	0.0607	0.0626	0.0504	0.0450	0.0623	0.0624	0.0448

Table 11 gives the prediction results. Comparing with the actual values in table 2, it is found that, except for the actual values of NO_x of sample 3 (fuel 10) and PM of the samples 1,2 (fuel 2,6) which are out of the intervals of the prediction results, the other actual values lie in the intervals of the prediction results. When the actual values do not lie in the intervals of the prediction results, they approach the upper limits or the lower limits of the prediction results, which is similar to that of H=0.8. Since the intervals of the prediction results are bigger than those of H=0.8, the prediction accuracy is worse than that of H=0.8. Comparing with table 8, it is also found that its prediction accuracy is better than that of H=0.5.

TABLE 11 THE PREDICTION RESULTS OF H=0.9 CALCULATED WITH THE FUZZY LINEAR REGRESSION EQUATIONS

The prediction samples		1	2	3
HC	Upper limits	0.0950	0.1203	0.0723
	Centers	0.0838	0.1017	0.0617
	Lower limits	0.0726	0.0830	0.0511
CO	Upper limits	0.4949	0.6320	0.3600
	Centers	0.4438	0.5291	0.3136
	Lower limits	0.3928	0.4262	0.2672
NO_x	Upper limits	0.5844	0.5727	0.5817
	Centers	0.5615	0.5498	0.5588
	Lower limits	0.5386	0.5268	0.5359
PM	Upper limits	0.0516	0.0657	0.0540
	Centers	0.0513	0.0654	0.0500
	Lower limits	0.0509	0.0651	0.0461

In conclusion, the prediction results of the equations for H = 0.8 are the best of the selected degrees of the membership, and those of H = 0.9 take second place, while those of H = 0.5 are the worst. When H is changed, the fuzzy centers of the coefficients of the equations for HC, CO and NO_x are not changed, but the fuzzy centers of the coefficients of the equations for PM are changed. Except when the fuzzy widths are equal to zero, the other fuzzy widths increase with H. Hence the intervals of the simulation and prediction results also increase.

Further Analysis and Discussion on the Restrictions

The simulation accuracy of the equations is decided by the fuzzy centers and fuzzy widths of the coefficients. The fuzzy centers decide the biases of the center values of the simulation results to the actual values. The fuzzy widths decide the sizes of the intervals of the simulation results. Hence, in order to make the biases small, an additional condition should be used to restrict the fuzzy center of the coefficients. In order to make the actual values lie in the intervals of the simulation results and the intervals not be too big to have meaning, an additional condition should also be used to restrict the fuzzy widths of the coefficients (Behrooz Heshmaty et al., 1985).

In terms of (7), the bias can be controlled by $\left|y_i - x_i{}^T\xi\right|$. It is denoted that with the ratio of $\left|y_i - x_i{}^T\xi\right|$ to y_i, the ratio is less than a percent selected by the decision-maker. It can also be considered as the form of the degree of the membership of the triangular membership function, i.e., $\mu_{y_i}(\left|x_i{}^T\xi - y_i\right|) \geq L$, where L is a value selected by the decision-maker in the L-set. That is to say:

$$\mu_{y_i}(\left|x_i{}^T\xi - y_i\right|) = 1 - \frac{\left|x_i{}^T\xi - y_i\right|}{y_i} \geq L \tag{16}$$

A similar idea can be applied to the fuzzy width $\mu_{y_i}(\sum_j \eta_j \left|x_{ij}\right| \geq M)$, here, M is a value also selected by the decision-maker in the M-set. The degree of the membership is also similarly defined with the triangular membership function, and the idea is shown with the following formula:

$$\mu_{y_i}(\sum_j \eta_j \left|x_{ij}\right|) = 1 - \frac{\sum_j \eta_j \left|x_{ij}\right|}{y_i} \geq M \tag{17}$$

Adding (16) and (17) into the restriction of the linear programming problem (8), a new linear programming problem is obtained. The fuzzy regression coefficients can be also obtained by the solution of the new linear programming problem.

However, in terms of (7), since the size of H is related to $\left|y_i - x_i{}^T\xi\right|$ and $\eta_j\left|x_{ij}\right|$, a bigger value of H may be corresponding to a less value of $\left|y_i - x_i{}^T\xi\right|$ and a bigger value of $\eta_j\left|x_{ij}\right|$. Only when the bias of $x_i{}^T\xi$ and y_i is big and $\eta_j\left|x_{ij}\right|$ closes or exceeds y_i, can satisfactory results be achieved by adding (16) and (17) into the restriction of (8).

In order to test the effect of the restrictions of (16) and (17) on the solution of the coefficients, H=0.8, L=0.95 and M=0.8 are selected and the coefficients of the equations are calculated. The coefficients of the equations of HC, CO and NO_x are the same as those of H=0.8, the coefficients of the equations of PM are different from those of H=0.8 and are shown in table 12. It shows that when H=0.8, the simulation accuracy of the equations of HC, CO and NO_x achieves or exceeds that of H=0.8, L=0.95 and M=0.8, so the restrictions do not operate. However, when calculating the coefficients of the equations of PM, the restrictions operate and different coefficients are obtained. Comparing the simulation results in table 13 with those in table 9, it is found that the biases of the center values of the simulation results to the actual values of PM for H=0.8 are bigger than those of H=0.8, L=0.95 and M=0.8, as well as the intervals of the simulation results are also bigger than those of H=0.8, L=0.95 and M=0.8, so the simulation accuracy of the equation for H=0.8, L=0.95 and M=0.8 is higher than that of H=0.8.

Table 12 THE FUZZY REGRESSION COEFFICIENTS OF H=0.8, L=0.95 AND M=0.8

		x_0	x_1	x_2	x_3	x_4
PM	ξ	0.0558386456	$4.74930849106\times10^{-4}$	$4.27655572740\times10^{-4}$	$2.60453547468\times10^{-4}$	$3.08841406342\times10^{-5}$
	η	0.0028701078	$4.45702491988\times10^{-6}$	0	$6.41944636672\times10^{-4}$	0

TABLE 13 THE SIMULATION RESULTS OF H=0.8, L=0.95 AND M=0.8 CALCULATED WITH THE FUZZY LINEAR REGRESSION EQUATIONS

The simulation samples		1	2	3	4	5	6	7	8
PM	Upper limits	0.0489	0.0618	0.0636	0.0512	0.0464	0.0662	0.0691	0.0509
	Centers	0.0481	0.0612	0.0630	0.0505	0.0461	0.0641	0.0656	0.0481
	Lower limits	0.0473	0.0606	0.0623	0.0498	0.0457	0.0620	0.0622	0.0454

Thus, an appropriate *H* must be selected to solve the coefficients. If the biases of the center values of the simulation results to the actual values are small and the intervals of the simulation results are much smaller than the actual values, the restrictions are not used; otherwise, the restrictions should be used to obtain the satisfactory results.

Conclusions

The coefficients of the independent variables obtained with the fuzzy linear regression are fuzzy numbers, so the results obtained with the equations are also fuzzy numbers.

For the solution of the coefficients is related to *H*, whose size is decided by the decision-maker. When the size of *H* is changed, the coefficients will be changed, and the simulation and prediction accuracy of the equations will also be changed. Thus, the appropriate *H* must be selected. In this paper, when H=0.8 is selected, the results are satisfactory.

When the biases of the center values of the simulation results to the actual values are big, or the intervals of the simulation results approach the actual values or are bigger than the actual values, the restrictions should be used to obtain the satisfactory results.

REFERENCES

Behrooz Heshmaty and Abraham Kandel, Fuzzy Linear Regression and Its Applications to Forecasting in Uncertain Environment. Fuzzy Sets and Systems 1985; 15: 151-191.

Hu G, Zhu R. Multiple data analysis method, Tianjing: Nankai University Press, 1990.

Hideo Tanaka, Satoru Uejima, and Kiyoji Asai. Linear Regression Analysis with Fuzzy Model. IEEE Transactions on System, Man, and Cybernetics 1982; 6: 903-907.

He J, Jiang D, Chen S. Practical linear programming and computer program, Beijing: Tsinghua University Press; 1985.

Li B, Zhu M, and Xu K. A practical engineering method for fuzzy reliability analysis of mechanical structures. Reliability Engineering and System Safety 2000; 67: 311-315.

Li H, Wang P. Fuzzy mathematics, Beijing: National defense industry press, 1994.

M.Hublin, P.G.Gadd, D.E.Hall, etc. European Programmes on Emissions, Fuels and Engine Technologies (EPEFE) — Light Duty Diesel Study, SAE961073. Journal of Fuels & Lubricants 1996; 4: 527-561.

Wu J. Device fault diagnosis using possibilistic theory. Fuzzy Systems and Mathematics 1999; 12: 25-32.